

MONTHLY WEATHER REVIEW

Editor, JAMES E. CASKEY, Jr.

Volume 78
Number 7

JULY 1950

Closed Sept. 5, 1950
Issued October 15, 1950

A NUMERICAL METHOD FOR FORECASTING RAINFALL IN THE LOS ANGELES AREA

J. C. THOMPSON¹

Weather Bureau Airport Station, Los Angeles, California
[Manuscript received June 16, 1950]

ABSTRACT

The application of modern statistical methods to the forecasting of rainfall in Los Angeles is discussed. Forecasts are made by graphical integration of a number of objective meteorological variables and the results presented in terms of the probability of rainfall occurring in each of several amount categories. The accuracy of this technique is discussed and compared with that obtained by current conventional forecasting methods, while the precision of the probability estimates is compared with a subjective evaluation of the probability distribution. Both comparisons show a slight, but statistically nonsignificant bias in favor of the numerical method.

The probability forecasts are shown to provide additional information regarding the reliability of each prediction which, by applying the principle of calculated risk, may be used to minimize the cost of carrying on any repetitive operation. An example of the use of this type of forecast is given, showing the saving which would result in a typical industrial operation in Los Angeles during the winter season of 1949-50.

CONTENTS

	Page
Abstract.....	113
Introduction.....	113
Forecast period.....	114
Method of development.....	114
Meteorological variables.....	115
Forecast accuracy.....	120
Comparison of objective and actual forecasts.....	121
Probability distribution.....	122
Use of probability forecasts.....	122
Conclusion.....	123
Acknowledgment.....	124
References.....	124

INTRODUCTION

As part of an extensive research program instituted by the U. S. Weather Bureau for the general purpose of improving the accuracy and usefulness of short-range weather forecasts, a study of the application of statistical methods to the forecasting of winter rainfall in Los Angeles was begun during the summer of 1945. Following

publication of the preliminary results a year later [1], a project to continue the investigation was established as a cooperative effort of the Weather Bureau and the University of California at Los Angeles. This investigation was continued for a two year period, covering a number of different methods of analysis, types of presentation and forecast periods [2, 3].

The study indicated that, for many industrial, agricultural, and military operations, some advantages are to be obtained from the use of numerical methods in weather forecasting. In general, these advantages ensue from the increased efficiency provided in preparation of the forecast, as well as from the fact that additional information may be made available by the inclusion of a reliability index for each prediction. Due to space restrictions, the following discussion does not cover the entire investigation, but only describes one of the most promising of the methods, shows how it was derived, indicates the nature and accuracy of the forecast it provides and, finally, gives an example of how the forecast may be used to best advantage.

¹Now in Short Range Forecast Development Section, U. S. Weather Bureau, Washington, D. C.

FORECAST PERIOD

The forecast of greatest public distribution and usefulness for the Los Angeles area is issued a short time before the beginning of the normal business day. It describes the weather to be expected during the time beginning at 1030 PST of the current day through 1630 PST of the following day, a period which is denoted in the forecast by the terms "this afternoon, tonight, and tomorrow." In view of the importance of this particular forecast the greater part of the research effort was directed at the development of an objective rainfall forecasting method which would cover this period. Data upon which the forecasts were based were limited to those available by the 0430 PST map time, and the results verified by the rainfall amounts which were observed at the Los Angeles Weather Bureau City Office during the period outlined above. The forecasting method covers the normal winter rainy season, from October 1 through March 31.

METHOD OF DEVELOPMENT

The forecasting method was developed by use of a graphical integration technique suggested by Brier [4]. Briefly, this process involves the selection of a number of independent (or as nearly independent as possible) meteorological variables which are believed to be related to the weather element to be forecast during some later period. The general procedure is then to work with the variables in pairs, each set of two variables being plotted on a scatter diagram with the independent variables as coordinates and the values of the dependent variate indicated beside each plotted point. In the case of a discrete two-valued variate, the independent variables are combined into a single derived variable by fitting a probability surface to the data, with the values of the probability isopleths used to express the functional relationship between the coordinate variables and the plotted variate. Where the element to be forecast is continuous, the variables are

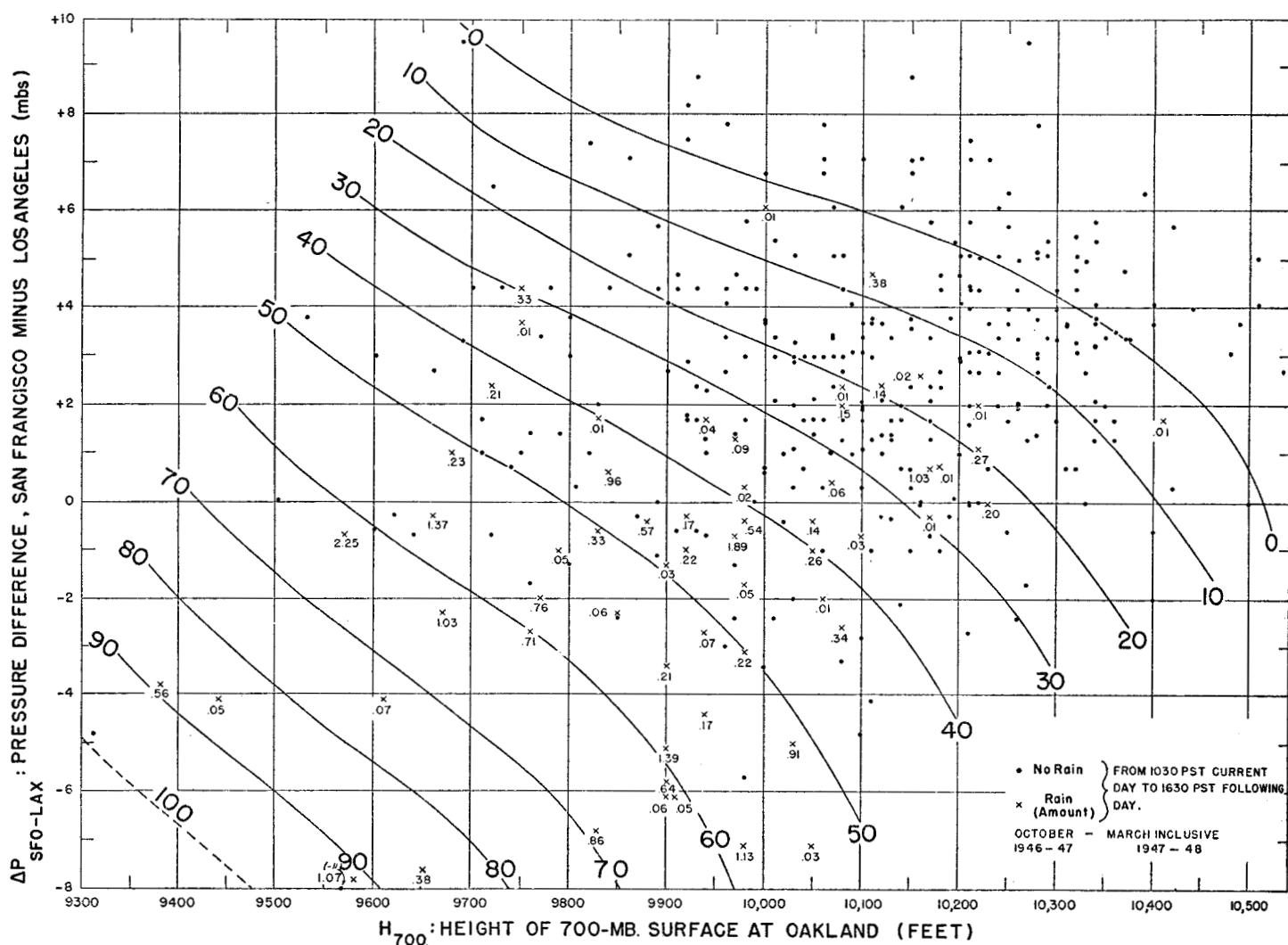


FIGURE 1.—Scatter diagram showing rainfall at Los Angeles as a function of $\Delta P_{SFO-LAX}$ and H_{700} . Solid curves are isograms of rainfall amount, adjusted to a scale of 0 to 100. These curves define a single variable X_1 , which is plotted as the abscissa in figure 4.

combined into a single derived variable by constructing isograms which express the values of the plotted variate.

The derived variables resulting from each pair of original variables are again combined in pairs and the process repeated until finally only one remains. This final derived variable is thus a function of all of the original variables and may accordingly be used to give some information about the weather element it is desired to forecast. A more complete description of the process is thought to be unnecessary here, since a fairly detailed discussion is available in Brier's original paper as well as in several recent forecasting studies by other investigators [5, 6, 7].

The graphical technique has the disadvantage of a certain amount of subjectivity in the original combination of variables, but this is largely outweighed by its relative simplicity as well as the fact that it eliminates the necessity for having prior knowledge of, or making assumptions regarding, the functional relationship between the independent variables and dependent variate, a requirement common to all mathematical regression methods. There is no lack of objectivity in the use of the charts obtained from the graphical analysis.

METEOROLOGICAL VARIABLES

A large number of meteorological variables which were

believed to be of significance in forecasting rainfall were tested, both singly and in various combinations. Here, use was made of the general approach to the problem of quantitative rainfall forecasting outlined by Showalter [8], as well as a particular application to major storms in the Los Angeles Basin presented in a report issued by the Hydrometeorological Section of the U. S. Weather Bureau [9]. Of the relationships tested, the following six variables produced the greatest skill and were incorporated in the final forecasting method:

H_{700}	Height of the 700-mb. surface at Oakland, vs
$\Delta P_{SFO-LAX}$	Sea level pressure difference, San Francisco minus Los Angeles.
P_{SFO}	Sea level pressure at San Francisco, vs
$\Delta P_{LAX-PHX}$	Sea level pressure difference, Los Angeles minus Phoenix.
D_{SDB}	Wind direction at Sandberg, vs
T_{700}	Temperature at 700 mb. at Santa Maria.

Although it is impractical to provide a complete discussion of the reasoning involved in testing of all variables investigated, a brief summary of the causal relationships which suggested the combination of the final six may be of interest.

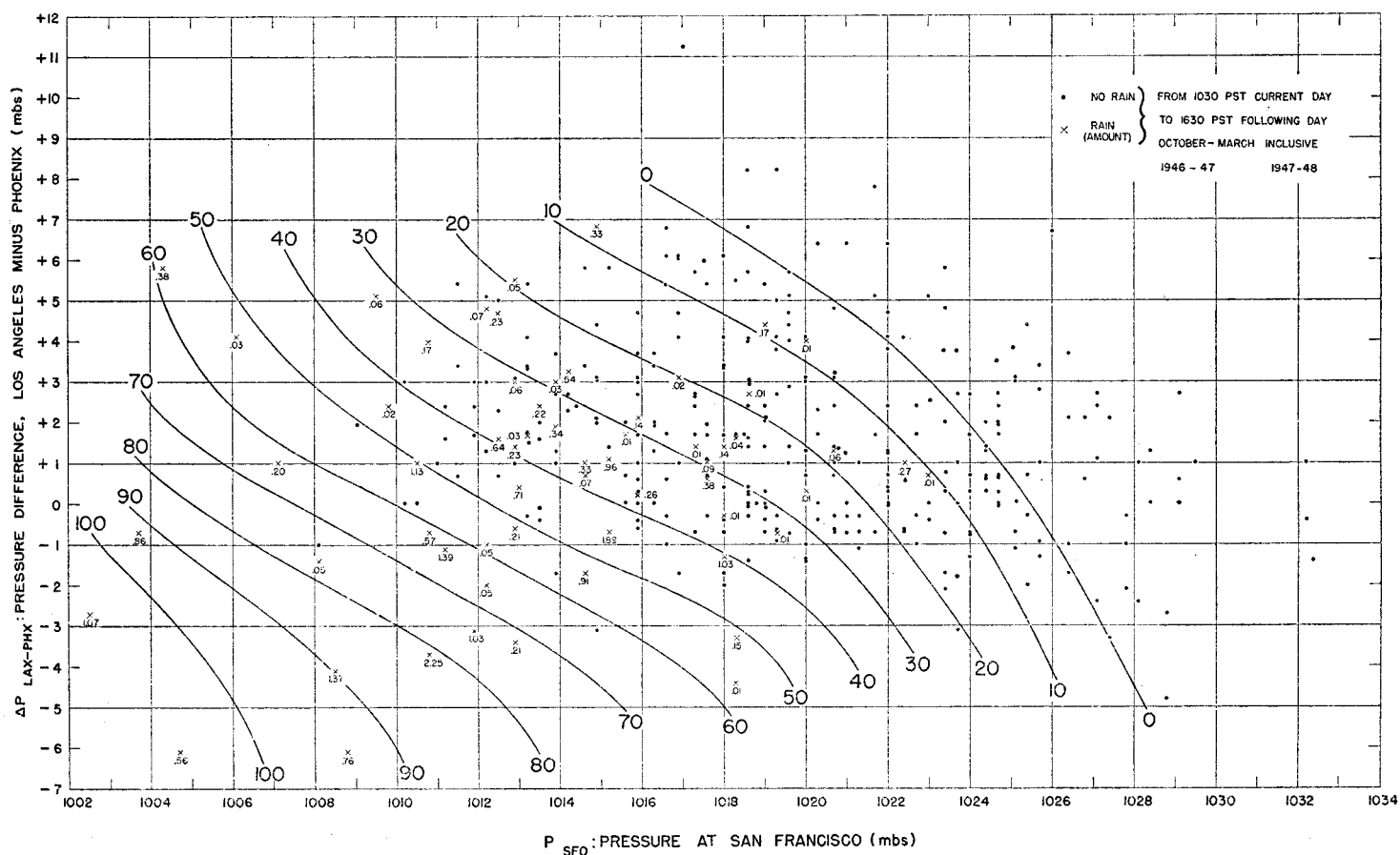


FIGURE 2.—Scatter diagram showing rainfall at Los Angeles as a function of $\Delta P_{LAX-PHX}$ and P_{SFO} . The solid curves, constructed as indicated in figure 1, define a variable X_2 , which is plotted as the ordinate in figure 4.

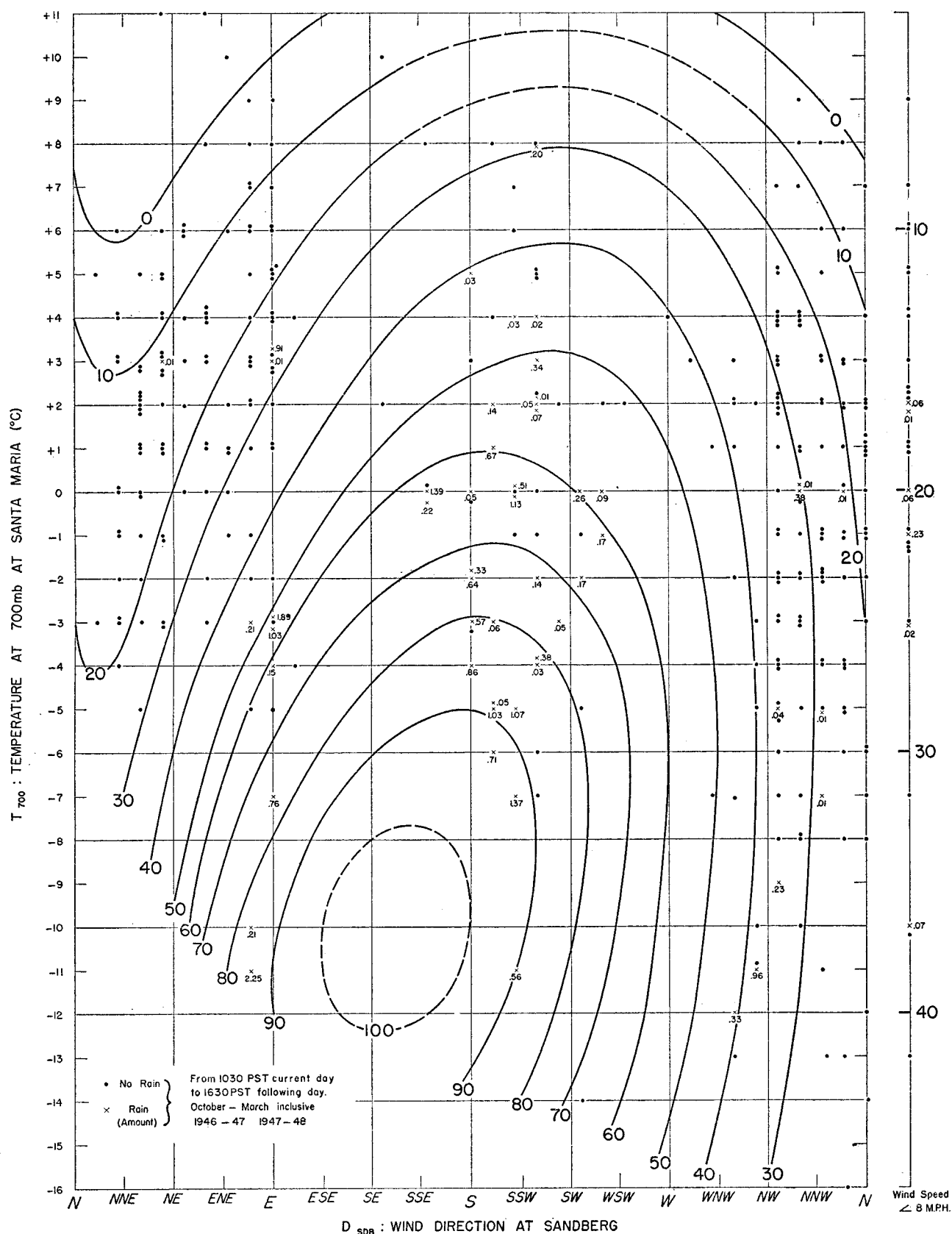


FIGURE 3.—Scatter diagram showing rainfall at Los Angeles as a function of T_{700} and D_{SD8} . The solid curves, constructed as indicated in figure 1, define a variable X_s , which is plotted as the abscissa in figure 5. At low wind speeds, the variable D_{SD8} loses its sensitivity as an indicator of the pressure field and thus was not used for speeds less than 8 m. p. h. These cases were plotted against T_{700} along the vertical axis to the right of the scatter diagram and analyzed separately to determine X_s .

It had previously been found that, for storms approaching from any westerly direction, the height of the 700-mb. surface at Oakland is quite well correlated with the distance of the nearest low center [2]. For storms of this type, the sea level pressure difference between Los Angeles and San Francisco, when combined with the distance of the low center, provides a rough measure of the strength of the wind flow across the area and indicates to some

extent the isobaric curvature, and hence horizontal convergence, associated with the disturbance. For storms approaching from the east, the above reasoning does not apply, but these are few in number and as a rule only small amounts of rain are associated with them.

When storms move inland a surface trough usually develops over the interior. Thus the sea level pressure difference, Los Angeles minus Phoenix, which is usually

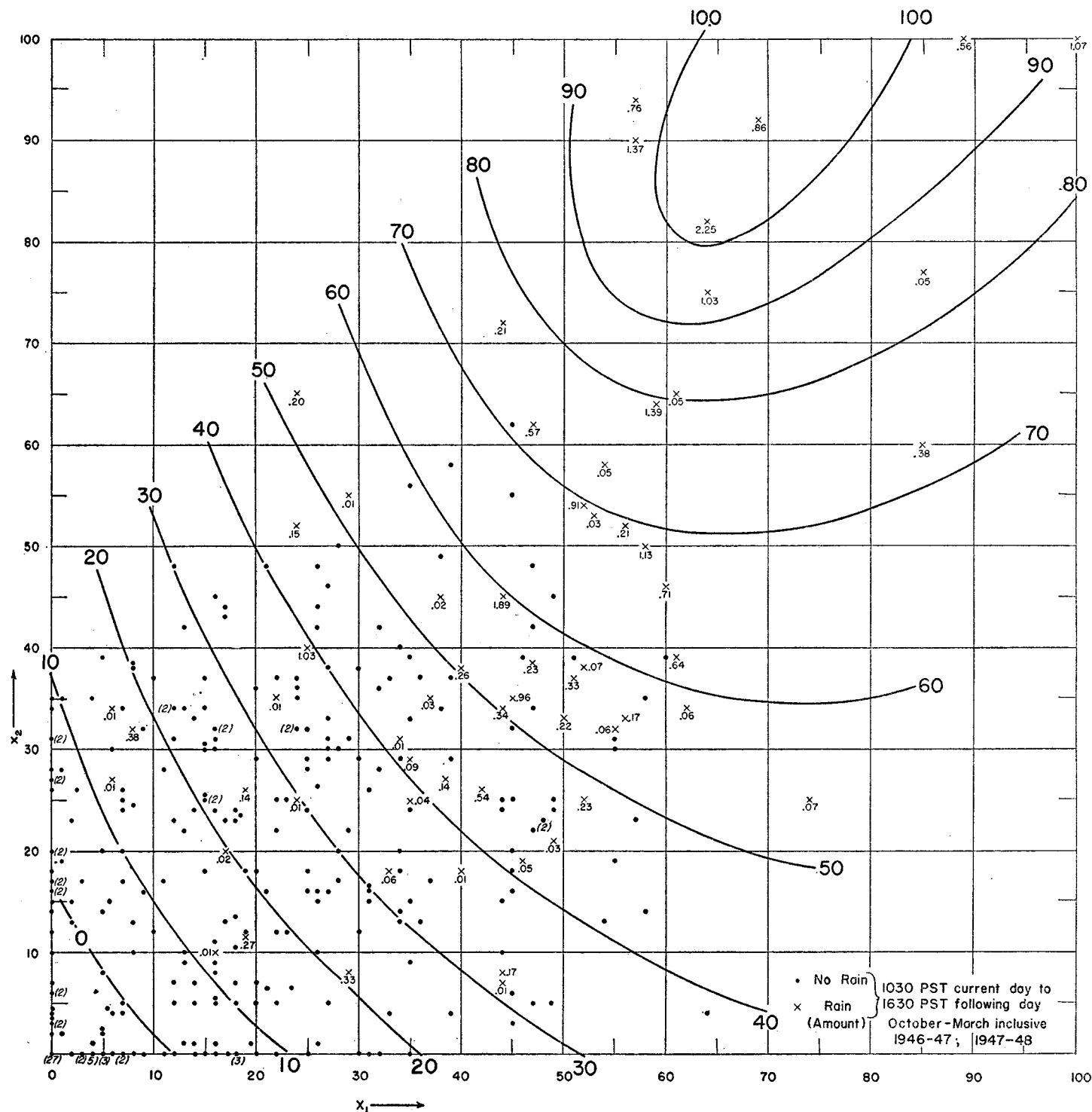


FIGURE 4.—Scatter diagram showing rainfall at Los Angeles as a function of X_1 (from fig. 1) and X_2 (from fig. 2). The solid curves, constructed as indicated in figure 1, define a variable Y_1 which is plotted as the ordinate in figure 5. The number in parentheses under a dot indicates the number of dots falling at that given point.

negative when a storm is situated off the coast, tends to become positive following, or just preceding, the end of the rain. This variable also helps to evaluate the rainfall resulting from easterly storms, which were not considered in the previous combination of variables. When combined with the sea level pressure at San Francisco, negative values of the pressure difference, Los Angeles minus Phoenix, and low values of the pressure at San Francisco are indicative of heavy rain.

The temperature at 700 mb. at Santa Maria provides a crude measure of the air mass stability, while the wind direction at Sandberg has long been used by experienced forecasters in this region as a rainfall forecasting aid to indicate the approach of a storm from the Pacific (southerly winds) or the final passage of a cold front (northerly winds). Winds at that point are apparently much more sensitive to changes in the pressure field than are those in the free air.

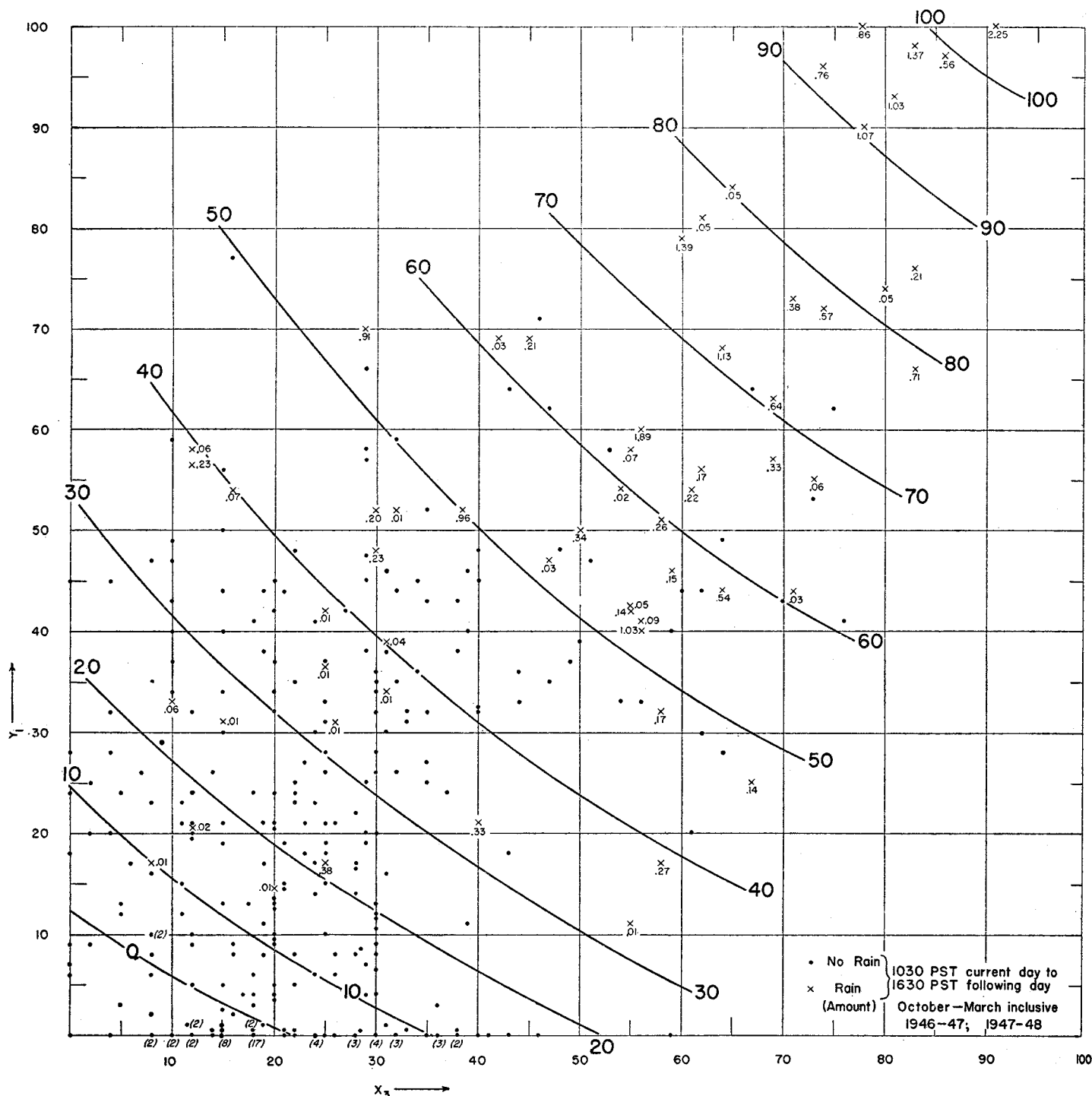


FIGURE 5.—Scatter diagram showing rainfall at Los Angeles as a function of X_3 (from fig. 3) and Y_1 (from fig. 4). The solid curves, constructed as indicated in figure 1, define variable Y_2 , which is plotted as the abscissa in figure 6. The number in parentheses under a dot indicates the number of dots falling at that given point.

While the meteorological relationships brought out by the primary graphical combination of each pair of variables (figs. 1, 2, and 3) may thus be discussed from a physical standpoint, and thereby the reasonableness of the isograms checked, very little can be said about the secondary combinations (figs. 4 and 5). Here the complexity of the joint relationships, as well as the probable effect of other variables not considered in the integration, defeats any attempt to supply a theoretical or physical justification for the distribution of the isograms. Consequently the construction of these charts must depend almost entirely upon an analysis of the data.

It will be noted that no measure of air mass moisture has been included in the system, since such moisture variables as were tested produced no significant increase in forecasting skill. This is probably due in part to the fact that dry air may prevail over this area up to within a short time of the beginning of rain. Furthermore, experience here more or less confirms the tentative conclusions of the Committee on Quantitative Rainfall Forecasting [10] that as a rule the kinematics of the cyclonic circulation (convergence and vertical motion) are much more important in determining rainfall intensity in this area than are variations in moisture.

Variables which are normally used to measure the velocity and/or deepening and filling of pressure systems, i. e., time derivatives of pressure or temperature, produced no apparent increase in forecasting skill. It should be noted, however, that since a measure of two components of the geostrophic wind is given by the pressure differences between San Francisco and Los Angeles, and Los Angeles and Phoenix, the addition of the actual wind at Sandberg provides a partial measure of the atmospheric accelerations which produce deepening and filling.

In general, no variable, or combination of variables was considered to have added information to the system unless its inclusion produced an increase in skill. This does not mean that variables other than those included in the integration are not related to the occurrence of rainfall in Los Angeles, but merely reflects the inability of the graphical analysis and/or analyst to supply the relationship, or indicates that these other variables are to a considerable degree correlated with those already used.

Figures 1 to 5, inclusive, show the result of the graphical integration process using the six variables listed above. The analysis of each chart was carried out by first constructing isograms of rainfall amounts and then adjusting the isogram values to a scale of 0 to 100. This latter device provides a uniform system of coordinates on all succeeding charts, which somewhat simplifies their preparation and final use by the forecaster. At the same time, the basis for the construction of the isograms, i. e., rainfall amounts, not probability of rainfall occurrence, is preserved.

Using the parameter Y_2 from figure 5 as the rainfall forecasting criterion, figure 6 was constructed, showing the

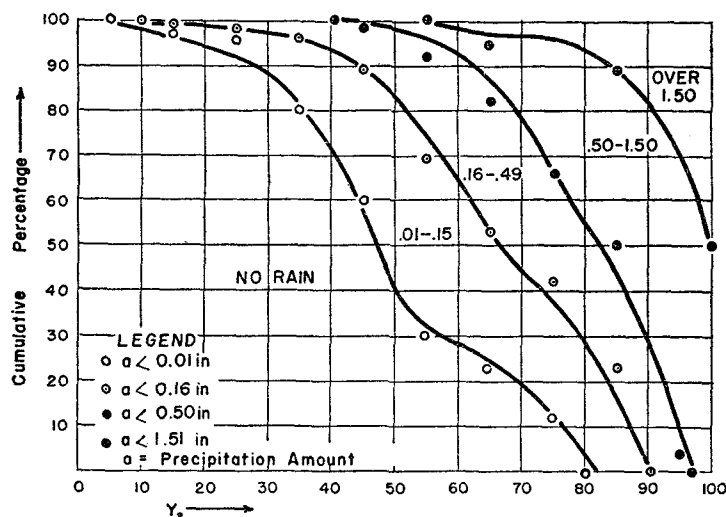


FIGURE 6.—Cumulative percentage frequency of selected rainfall amount categories as a function of the final variable Y_2 . Rainfall probability values obtained from this graph are given in table 1.

percentage frequency of rainfall occurring in each of five amount categories. These categories were selected in order to agree with those used in the official verification of rainfall forecasts (PFR system). From the smoothed curves shown in figure 6, table 1 was prepared.

TABLE 1.—Relation between Y_2 and the probability that rain will occur in the indicated categories

Y_2	No rain	0.01-0.15	0.16-0.49	0.50-1.50	Y_2	No rain	0.01-0.15	0.16-0.49	0.50-1.50	1.51 or more
0	100				51	38	44	15	3	
1	100				52	37	43	16	4	
2	100				53	35	44	17	4	
3	100				54	34	43	18	5	
4	100				55	32	43	20	5	
5	100				56	32	42	20	5	1
6	100				57	31	41	22	5	1
7	100				58	30	40	23	6	1
8	99	1			59	29	38	25	6	2
9	99	1			60	28	36	28	6	2
10	98	2			61	28	34	29	7	2
11	98	2			62	27	32	31	7	3
12	98	2			63	26	31	33	7	3
13	97	3			64	25	30	33	9	3
14	97	3			65	25	27	35	10	3
15	96	4			66	24	26	36	11	3
16	96	4			67	23	25	36	13	3
17	96	3	1		68	22	25	35	14	4
18	95	4	1		69	21	25	34	16	4
19	94	5	1		70	20	25	34	17	4
20	94	4	2		71	19	25	32	20	4
21	94	4	2		72	17	25	32	22	4
22	93	5	2		73	16	25	31	24	4
23	93	5	2		74	14	25	31	25	5
24	92	6	2		75	13	25	29	28	5
25	92	6	2		76	11	25	28	31	5
26	91	7	2		77	10	25	27	32	6
27	90	8	2		78	8	24	28	34	6
28	90	7	3		79	6	24	28	36	6
29	89	8	3		80	4	24	27	38	7
30	88	9	3		81	2	24	27	40	7
31	86	11	3		82		24	27	41	8
32	85	12	3		83		22	26	43	9
33	84	12	4		84		20	26	44	10
34	83	13	4		85		16	26	47	11
35	81	15	4		86		13	26	49	12
36	80	16	4		87		9	26	51	14
37	78	17	5		88		6	26	52	16
38	76	18	6		89		2	26	54	18
39	74	20	6		90			25	55	20
40	72	21	7		91			22	56	22
41	69	23	8		92			18	58	24
42	66	25	8	1	93			13	61	26
43	64	26	9	1	94			9	62	29
44	60	30	8	2	95			4	64	32
45	58	31	9	2	96				64	36
46	54	34	10	2	97				60	40
47	50	36	12	2	98				57	43
48	47	38	13	2	99				51	49
49	44	40	13	3	100				48	52
50	41	42	14	3						

In order to facilitate practical use of the method, mimeographed work sheets containing a schematic diagram of the combination process, as well as other pertinent information, were prepared. Figure 7 illustrates a portion of this work sheet, and figures 8 and 9 are examples of forecasts made using the method.

FORECAST ACCURACY

Tables 2, 3, and 4 are contingency tables showing forecast and observed precipitation amounts for original data, independent test data, and actual forecasts made at the Los Angeles Forecast Center during the same period as the test data. Objective forecasts for both original and test data were made from computed values of Y_2 as shown in the table in the work sheet, figure 7.

TABLE 2.—Contingency table showing verification of objective forecasts for original data (October–March 1946–47 and 1947–48)

Observed precipitation (inches)	Forecast precipitation (inches)						Total
	No rain	0.01–0.15	0.16–0.49	0.50–1.50	Over 1.50		
No rain	280	13	5	0	0		298
0.01–0.15	15	9	1	3	0		28
0.16–0.49	7	2	4	2	0		15
0.50–1.50	0	4	2	9	0		15
Over 1.50	0	0	1	0	1		2
Total	302	28	13	14	1		358

Percentage correct: $\frac{303}{358} = 0.85$; Skill score*: $\frac{303-255}{358-255} = 0.47$

*The skill score is defined as the ratio of correct forecasts to total forecasts exceeding those which would be correct if the same forecasts were distributed by chance. The formula for the skill score (S) is thus:

$$S = \frac{C - E}{T - E}$$

where C = number of correct forecasts, T = total number of forecasts, E = number of forecasts expected correct due to chance. The theory and procedure involved in computation of E may be found in almost any statistical text.

Weather Bureau Airport Station
Los Angeles, Calif.

OBJECTIVE PRECIPITATION FORECAST FOR PERIOD 6 TO 36 HOURS FOLLOWING MAP TIME

Date _____
Map time _____ PST

$\Delta P_{SFO-LAX}$ _____	} (fig. 1) X_1 _____	} (fig. 5) Y_2 _____
H_{700} _____		
$\Delta P_{LAX-PHX}$ _____	} (fig. 2) X_2 _____	
P_{SFO} _____		
T_{700} _____	} (fig. 3) X_3 _____	
D_{SDB} _____		

When Y_2 is:	Most probable amount of rain to be forecast is:	Relative probability, % (from table 1)
0 - 49	No rain	
50 - 62	.01 - .15	
63 - 75	.16 - .49	
76 - 99	.50 - 1.50	
100	1.51 or more	

FIGURE 7.—Sample of forecasting work sheet showing method of combining variables, and giving most probable rain amounts to be forecast for categorical values of Y_2 . See page 115 for definitions of symbols.

TABLE 3.—Contingency table showing verification of objective forecasts for independent data (October–March 1944–45 and 1945–46)

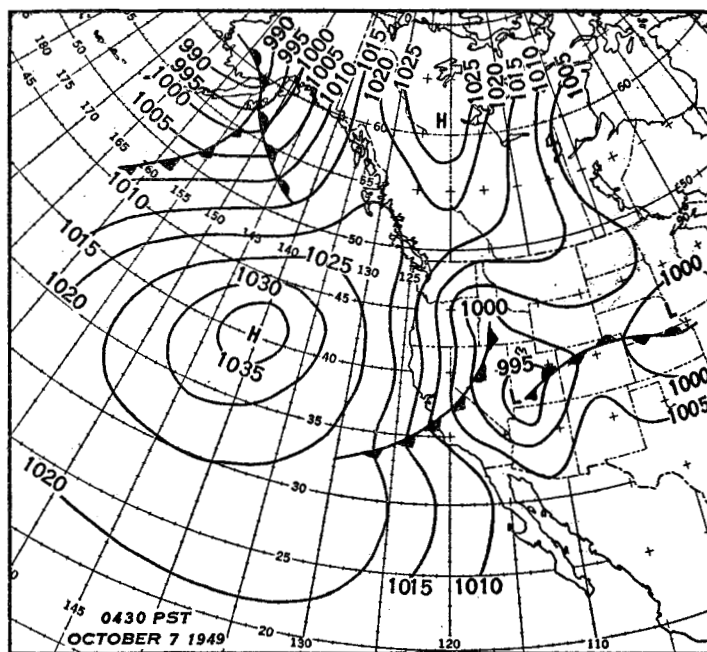
Observed precipitation (inches)	Forecast precipitation (inches)						Total
	No rain	0.01–0.15	0.16–0.49	0.50–1.50	Over 1.50		
No rain	273	11	5	1	0		290
0.01–0.15	16	4	6	3	0		29
0.16–0.49	3	4	7	1	0		15
0.50–1.50	6	2	4	9	0		21
Over 1.50	0	0	1	0	0		1
Total	298	21	23	14	0		356

Percentage correct: $\frac{293}{356} = 0.82$; Skill score: $\frac{293-246}{356-246} = 0.43$

TABLE 4.—Contingency table showing verification of actual forecasts, PFR system (October–March 1944–45 and 1945–46)

Observed precipitation (cumulative code number)	Forecast precipitation (cumulative code number)						Total
	0	1	2	3	≥4		
0	244	15	4	4	2		269
1	15	6	4	2	2		29
2	8	3	2	0	2		15
3	2	3	2	2	7		16
≥4	2	4	3	8	18		35
Total	271	31	15	16	31		364

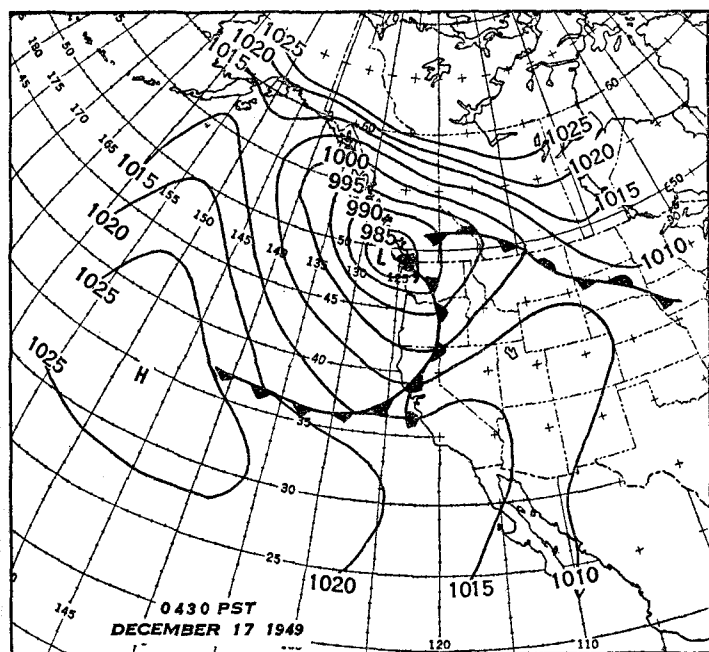
Percentage correct: $\frac{272}{364} = 0.75$; Skill score: $\frac{272-207}{364-207} = 0.41$



The PFR forecast verifications (table 4) are for the same period and, except for code number "1," which at that time included rainfall of a trace to 0.15 inch, instead of the interval 0.01 to 0.15 inch, the code numbers cover the same categories as the independent forecasts in the previous table. However, it should be noted that the PFR forecasts were made for three separate time intervals during the forecast period; i. e., this afternoon, tonight, and tomorrow. In the above table, the code numbers (0 to 4) for the three time intervals were simply added and verifications were based on these cumulated values. Since it is difficult to say whether or not either of the forecasting methods might have been favored by this procedure, the justification for comparing the independent objective forecast and the actual PFR forecast scores is somewhat doubtful.

COMPARISON OF OBJECTIVE AND ACTUAL FORECASTS

In order to make a more valid comparison of the



$\Delta P_{\text{SFO-LAX}}$	<u>-6.1</u>	}	X_1	<u>69</u>
H_{700}	<u>9790</u>			
$\Delta P_{\text{LAX-PHX}}$	<u>0.0</u>	}	X_2	<u>60</u>
P_{SFO}	<u>1009.8</u>			
		}	T_{700}	<u>0</u>
		}	$D_{\text{SDB SSW}}$	<u> </u>

<div style="font-size: 3em; padding: 0 10px;">}</div> <div style="padding: 5px;">Y_1</div> <div style="text-align: right; padding: 5px;"><u>76</u></div>		Relative Probability (From table I)	
		No Rain	.11
		.01 - .15	.25
		.16 - .49	.28
		.50 - 1.50	.31
		Over 1.50	.05

MOST PROBABLE AMOUNT FORECAST: .50 - 1.50 in.

ACTUAL RAINFALL: 1.15 in.

FIGURE 9.—Example of surface weather map and objective forecast for frontal situation which produced over an inch of rain at Los Angeles.

objective and actual forecasts, as well as to determine, if possible, whether or not conventional forecasting methods are able to add significantly to the accuracy of the objective system, a comprehensive test was arranged for the winter (October–March) of 1949–50. Two forecasts were made each day, by two different forecasters, the first being made at 0700 PST and the second at 0800 PST. In the first instance, the forecaster had available all information used in the objective system including the computed objective forecast and analyzed 0430 PST surface map. The second forecaster had the added advantage of being able to check on the data for the following three-hourly (0730 PST) surface chart.

Forecasts were made and verified for the same rainfall categories, and for the same period as the objective system. Furthermore, in order to minimize the areal variation produced by a single station rainfall measurement, the amounts were verified by using unweighted means of the precipitation recorded at three weather stations in the Los Angeles Basin; i. e., Los Angeles Airport, Los Angeles City Office, and Burbank Airport.

Results of this test are given in tables 5, 6, and 7.

Although the skill shown by the objective forecast is greater than for either of the other forecasts, a statistical (Chi-square) test indicates that, at the 5 percent significance level, the differences in frequency distributions for the three contingency tables may be assumed due to chance variations. This means that there is no significant difference in the accuracy of the three forecasts,

TABLE 5.—Contingency table showing verification of objective forecasts
(Oct. 1, 1949—March 31, 1950)

		Forecast precipitation (inches)					
		No rain	0.01-0.15	0.16-0.49	0.50-1.50	Over 1.50	Total
Observed precipitation (inches)	No rain	145	4	2	0	0	151
	0.01-0.15	6	3	2	2	1	14
	0.16-0.49	3	0	4	0	0	7
	0.50-1.50	2	0	4	3	0	9
	Over 1.50	0	0	1	0	0	1
Total		156	7	13	5	1	*182

Percentage correct: $\frac{155}{182} = 0.85$; Skill score: $\frac{155 - 131}{182 - 131} = 0.47$

*Differences in total forecasts in Tables 5, 6, and 7 are due to a few missing data in the latter two tabulations. However, a computation of percentage of correct forecasts and skill scores for those days for which information was available for all three forecasts resulted in exactly the same scores as those listed.

TABLE 6.—Contingency table showing verification of actual forecasts made at 0700 PST (Oct. 1, 1949—March 31, 1950)

		Forecast precipitation (inches)					
		No rain	0.01-0.15	0.16-0.49	0.50-1.50	Over 1.50	Total
Observed precipitation (inches)	No rain.....	140	9	0	0	0	149
	0.01-0.15.....	7	2	2	2	1	14
	0.16-0.49.....	2	2	2	1	0	7
	0.50-1.50.....	3	2	1	2	0	8
	Over 1.50.....	0	0	0	1	0	1
	Total.....	152	15	5	6	1	179

Percentage correct: $\frac{146}{179} = 0.82$; Skill score: $\frac{146-123}{179-123} = 0.35$

TABLE 7.—Contingency table showing verification of actual forecasts made at 0800 PST (Oct. 1, 1949–Mar. 31, 1950)

Observed precipitation (inches)	Forecast precipitation (inches)					Total
	No. rain	0.01–0.15	0.16–0.49	0.50–1.50	Over 1.50	
No rain	142	8	0	0	0	150
0.01–0.15	8	1	4	1	0	14
0.16–0.49	3	2	1	1	0	7
0.50–1.50	1	0	5	3	0	9
Over 1.50	0	0	1	0	0	1
Total	154	11	11	5	0	181

Percentage correct: $\frac{147}{181}=0.81$; Skill score: $\frac{147-129}{181-129}=0.35$

and confirms the previous comparison made between the independent test sample, table 3, and the actual PFR forecasts, table 4.

Any attempt to generalize on the above results in the light of their possible effect on current forecasting procedures is beyond the scope of this discussion. Here it is desired only to suggest that, in this case at least, the numerical forecasting technique produced results which were at least as accurate as, and were not improved upon by, conventional methods. At the same time, by presenting the forecast in terms of probabilities, the numerical method provided a measure of the reliability of each prediction.

PROBABILITY DISTRIBUTION

Occasional attempts have been made to provide probability forecasts in the past, notably by Besson [11] in France, Cooke [12] in Australia, and Hallenbeck [13] in the United States. Except in the case of Besson's studies, however, all such forecasts were based upon subjective estimates of the probability distribution and were consequently dependent upon the individual forecasters' experience, skill, and certain psychological factors. The numerical forecasts discussed here are not subject to such influences and, at the same time, are apparently quite as accurate as those issued by conventional methods.

It should be noted, however, that the accuracy of the categorical forecasts may not necessarily reflect the precision of the probability estimates. A method for evaluating the latter, suggested recently by G. W. Brier [14], is described briefly below. If the probability estimates are placed in a contingency table as follows,

	Forecasts		
	1	2	...n
Forecast events	1	p_{11}	p_{12} p_{1n}
	2	p_{21}	p_{22} p_{2n}
	...r	p_{r1}	p_{r2} p_{rn}

where the p_{ij} are the forecast probabilities in the i th row and j th column, then the reliability of the forecasts (P) may be defined as,

$$P = \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^n (p_{ij} - E_{ij})^2 \quad (1)$$

where E_{ij} is 0 when the forecast event does not occur, and E_{ij} is 1 when the forecast event does occur.

Here the E_{ij} are the actual, or observed, probabilities so essentially what is done in the above formula is to compute the mean of the squares of the differences between the forecast probability distribution and the observed distribution. If the forecast events are mutually exclusive, the reliability score has a range of from zero to two. Since one would like to have the difference as small as possible, a good score is one which is small.

A rough check on the consistency of probability forecasts made by several individual forecasters in competition with the objective system was carried on at Los Angeles for a short period during the past winter. Due to schedule differences, days off, etc., forecasters' probability estimates were not made every day; consequently the comparison has been made between forecasters' scores and objective scores for only those days on which a forecast was made by the former. Results of this comparison are given in table 8.

TABLE 8.—Comparison of probability forecasts made by forecasters and objective method. Objective scores are computed for days on which forecasts were made by the indicated forecaster

Forecaster	Number of forecasts	Reliability score (P)	
		Forecaster	Objective
A	75	0.26	0.27
B	71	.35	.32
C	72	.35	.29
D	40	.39	.42
Mean	65	.34	.31

Here the average score for the group is slightly higher (and thus worse) than that for the objective method, while individual scores range from 0.03 lower to 0.06 higher than the objective system. Furthermore, an inspection of the individual forecasts reveals considerable variability between forecasters in the distribution of probability estimates on most days when rain is likely. It would therefore appear that in this case a more reliable estimate of the error frequency distribution for each forecast may be obtained from the objective method.

USE OF PROBABILITY FORECASTS

The usefulness of the reliability measurement provided by the probability forecast may be brought out by applying the well-known principle of calculated risk. As is true of statistical techniques in general, this principle requires that the decisions based on the forecasts be applied only to repetitive operations. Furthermore, the user in this case should have available complete information concerning the cost of each operation as well as an estimate of the

contingent gain or loss which will result if the forecast events do not occur. Then, in order to keep the cost of the series of operations at a minimum, decisions should be made by balancing the probability of occurrence of the forecast event against the ratio of the cost to the contingent gain or loss. This means that the usual categorical forecast may not be the most valuable prediction for all recipients since it is aimed at providing a forecast to suit the "average" user and is quite properly based (either subjectively or objectively) on the probability of occurrence being greater or less than 0.50.

Some uses of probability forecasts have already been discussed briefly by Brier [15] and Price [16] for certain special types of weather problems. However, it may be of interest to provide an example of how the rainfall forecasts described herein could be used to good advantage. Consider, therefore, a hypothetical Los Angeles construction company which is engaged in making a series of concrete pours during the winter months. The company finds that it will cost about \$400 to protect the concrete each time, but that damage of \$5,000 will result if rainfall exceeding 0.15 inch occurs within 36 hours of the time of pouring. Accordingly, for the cost of the entire operation to be minimized, the concrete should be protected if the ratio of the cost (C) to the contingent loss (L) is less than the probability (P) of rainfall greater than 0.15 inch. Thus, in this case,

$$P_{(\text{rain} > 0.15)} > \frac{C}{L} = \frac{400}{5000} = 0.08 \quad (2)$$

Consulting table 1, it will be seen that this inequality will be satisfied for any value of Y_2 exceeding 41. However, if the usual categorical forecast were used as the basis for this operation, the decision would be based on the probability being greater than 0.50, or,

$$P_{(\text{rain} > 0.15)} > .50 \quad (3)$$

This inequality is satisfied for any value of Y_2 exceeding 66. Accepting this as a basis for his decision, it is apparent that the contractor would not protect his concrete often enough.

In order to make this point clear, the comparison given below has been made of the cost of carrying on this operation throughout the past winter season (October 1, 1949–March 31, 1950) for several alternative procedures. Here, for the sake of simplicity, it is assumed that the contractor operates every day during the period (182 days).

Case I.—The contractor obtains no forecast at all and takes no protective measures. Since there were 17 days with rain exceeding 0.15 inch during the period, the loss is \$5,000 per day for 17 days----- \$85,000

Case II.—The contractor obtains no forecast but takes protective measures every day. He thus sustains no loss, but the cost is \$400 per day for 182 days----- 72,800

Case III.—The contractor uses climatological expectancy as a basis for the operation, taking protective measures only during the period when this expectancy is greater than 0.08. For Los Angeles this requires protection from December 11 until March 25 and no protection before or after that period. In this case the contractor would take protective measures on 105 days and would sustain a loss on 4 days----- \$62,000

Case IV.—The contractor uses "persistence"; i. e., takes protective measures on all days following a day with measurable rainfall and provides no protection on other days. This would require protection on 31 days and the contractor would suffer loss on 5 days----- 37,400

Case V.—The contractor obtains a forecast designed for the "average" user and thus based on equation (3) above. Using the objective probability estimates, this would require protective measures on 19 days, and the contractor would suffer loss on 5 days----- 32,600

Case VI.—The contractor obtains a forecast designed for his particular operation and thus based on equation (2). Using the objective probability estimates, this would require protective measures on 35 days, and the contractor would suffer loss on 2 days----- 24,400

Total cost
plus loss

An inspection of these figures reveals that, although the contractor in using the "average" forecast, Case V, would reduce the total cost of the operation below that for an operator using no forecast at all, or using climatological expectancy or persistence, the least total expenditure would result from the use of the probability forecast. This illustrates the advantage of the probability estimate and reveals the inherent danger in any categorical forecast where the user is not provided with, or does not make use of a measure of the reliability of the prediction. In the above example, only the objective probability forecasts have been considered, although probability estimates might also be made by the forecaster from a subjective evaluation of the meteorological situation. For some purposes where adequate numerical techniques are not available, such subjective probability forecasts undoubtedly could be used to good advantage.

CONCLUSION

The gradual increase in the complexity of modern industrial, agricultural, military, and many other operations has resulted, during recent years, in a general desire for more accurate and increasingly specialized weather forecasts. The success with which such requirements can be met is of course dependent largely upon basic progress in the science of meteorology in general. At present, however, many of the physical relationships involved in weather forecasting are obscure, and of those that are understood a large number are mathematically indeterminate. Whether or not statistical techniques may be used to help in the development of a better physical understanding of the weather is beyond the scope of this discussion. It is desired here only to point out the usefulness of such techniques in evaluating the magnitude of the indeterminacy, and to suggest a method for making use of that evaluation in a practical application.

ACKNOWLEDGMENT

Much of the research work involved in this study was carried on in cooperation with the Department of Meteorology, University of California at Los Angeles and was under the joint supervision of Prof. Jacob Bjerknes, Chairman of the Department, and the writer. Research assistants at U. C. L. A. were R. Robert Rapp, David Smedley, J. K. Angell, and Mrs. C. K. Chen-Wei, and at the Los Angeles Forecast Center was Mrs. Gwendolyn McMeans. Many valuable suggestions were also contributed by Mr. A. K. Showalter and the forecast staff at the Los Angeles Forecast Center.

REFERENCES

1. J. C. Thompson, "Progress Report on an Objective Rainfall Forecasting Research Program for the Los Angeles Area," U. S. Weather Bureau *Research Paper No. 25*, July 1946.
2. J. C. Thompson, R. R. Rapp and D. Smedley, *Second Progress Report on an Objective Rainfall Forecasting Program for the Los Angeles Area*, Dept. of Meteorology, University of California at Los Angeles, July 1947.
3. J. K. Angell and C. K. Chen, *Final Report on Objective Rainfall Forecasting Program for the Los Angeles Area*, Dept. of Meteorology, University of California at Los Angeles, May 1948.
4. G. W. Brier, "A Study of Quantitative Precipitation Forecasting in the T. V. A. Basin," U. S. Weather Bureau *Research Paper No. 26*, November 1946.
5. Samuel Penn, "An Objective Method for Forecasting Precipitation Amounts from Winter Coastal Storms for Boston," *Monthly Weather Review*, vol. 76, No. 8, August 1948, p. 149-161.
6. D. L. Jorgensen, "An Objective Method of Forecasting Rain in Central California During the Raising and Drying Season," *Monthly Weather Review*, vol. 77, No. 2, February 1949, p. 31-46.
7. W. W. Dickey, "Estimating the Probability of a Large Fall in Temperature at Washington, D. C.," *Monthly Weather Review*, vol. 77, No. 3, March 1949, p. 67-78.
8. A. K. Showalter, "An Approach to Quantitative Forecasting of Precipitation," *Bulletin of the American Meteorological Society*, vol. 25, No. 4, April 1944, p. 137-142; No. 7, September 1944, p. 276-288.
9. U. S. Weather Bureau, "Revised Report on Maximum Possible Precipitation, Los Angeles Area, California," *Hydrometeorological Report No. 21B*, December 29, 1945.
10. A. K. Showalter, A. J. Knarr, and R. D. Fletcher, Committee Report on Quantitative Rainfall Forecasting, U. S. Weather Bureau, August 6, 1948. (Unpublished.)
11. Louis Besson, "Essai de Prevision Methodique du Temps," *Annales de L'Observatoire Municipal, Ville de Paris*, Tome VI, 1905, 473-495.
12. W. E. Cooke, "Forecasts and Verifications in Western Australia," *Monthly Weather Review*, vol. 34, No. 1, January 1906, p. 23-24.
13. C. Hallenbeck, "Forecasting Precipitation in Percentages of Probability," *Monthly Weather Review*, vol. 48, No. 11, November 1920, p. 645-647.
14. G. W. Brier, "Verification of Forecasts Expressed in Terms of Probability," *Monthly Weather Review*, vol. 78, No. 1, January 1950, p. 1-3.
15. G. W. Brier, "Verification of a Forecaster's Confidence and the Use of Probability Statements in Weather Forecasting," U. S. Weather Bureau *Research Paper No. 16*, February 1944.
16. Saul Price, "Thunderstorm Today?—Try a Probability Forecast," *Weatherwise*, vol. 2, No. 3, June 1949, p. 61-63.